

7

A Roadmap for Revision

In reviewing the draft assessment *Toxicological Review of Formaldehyde-Inhalation Assessment: In Support of Summary Information on the Integrated Risk Information System (IRIS)*, the committee initially evaluated the general methodology (Chapter 2) and then considered the dosimetry and toxicology of formaldehyde (Chapter 3) and the review of the evidence and selection of studies related to noncancer and cancer outcomes (Chapters 4 and 5). Finally, the committee addressed the calculation of the reference concentrations (RfCs) for noncancer effects and the unit risks for cancer and the treatment of uncertainty and variability (Chapter 6). In this chapter, the committee provides general recommendations for changes that are needed to bring the draft to closure. On the basis of “lessons learned” from the formaldehyde assessment, the committee offers some suggestions for improvements in the IRIS development process that might help the Environmental Protection Agency (EPA) if it decides to modify the process. As noted in Chapter 2, the committee distinguishes between the process used to generate the draft IRIS assessment (that is, the development process) and the overall process that includes the multiple layers of review. The committee is focused on the development of the draft IRIS assessment.

CRITICAL REVISIONS OF THE CURRENT DRAFT IRIS ASSESSMENT OF FORMALDEHYDE

The formaldehyde draft IRIS assessment has been under development for more than a decade (see Chapter 1, Figure 1-3), and its completion is awaited by diverse stakeholders. Here, the committee offers general recommendations—in addition to its specific recommendations in Chapters 3-6—for the revisions that are most critical for bringing the document to closure. Although the committee suggests addressing some of the fundamental aspects of the approach to generating the draft assessment later in this chapter, it is not recommending that the assessment for formaldehyde await the possible development of a revised ap-

proach. The following recommendations are viewed as critical overall changes needed to complete the draft IRIS assessment:

- To enhance the clarity of the document, the draft IRIS assessment needs rigorous editing to reduce the volume of text substantially and address redundancy and inconsistency. Long descriptions of particular studies, for example, should be replaced with informative evidence tables. When study details are appropriate, they could be provided in appendixes.
- Chapter 1 needs to be expanded to describe more fully the methods of the assessment, including a description of search strategies used to identify studies with the exclusion and inclusion criteria clearly articulated and a better description of the outcomes of the searches (a model for displaying the results of literature searches is provided later in this chapter) and clear descriptions of the weight-of-evidence approaches used for the various noncancer outcomes. The committee emphasizes that it is not recommending the addition of long descriptions of EPA guidelines to the introduction, but rather clear concise statements of criteria used to exclude, include, and advance studies for derivation of the RfCs and unit risk estimates.
- Standardized evidence tables for all health outcomes need to be developed. If there were appropriate tables, long text descriptions of studies could be moved to an appendix or deleted.
- All critical studies need to be thoroughly evaluated with standardized approaches that are clearly formulated and based on the type of research, for example, observational epidemiologic or animal bioassays. The findings of the reviews might be presented in tables to ensure transparency. The present chapter provides general guidance on approaches to reviewing the critical types of evidence.
- The rationales for the selection of the studies that are advanced for consideration in calculating the RfCs and unit risks need to be expanded. All candidate RfCs should be evaluated together with the aid of graphic displays that incorporate selected information on attributes relevant to the database.
- Strengthened, more integrative, and more transparent discussions of weight of evidence are needed. The discussions would benefit from more rigorous and systematic coverage of the various determinants of weight of evidence, such as consistency.

FUTURE ASSESSMENTS AND THE IRIS PROCESS

This committee's review of the draft IRIS assessment of formaldehyde identified both specific and general limitations of the document that need to be addressed through revision. The persistence of limitations of the IRIS assessment methods and reports is of concern, particularly in light of the continued evolution of risk-assessment methods and the growing societal and legislative pressure to evaluate many more chemicals in an expedient manner. Multiple

groups have recently voiced suggestions for improving the process. The seminal “Red Book,” the National Research Council (NRC) report *Risk Assessment in the Federal Government: Managing the Process*, was published in 1983 (NRC 1983). That report provided the still-used four-element framework for risk assessment: hazard identification, dose-response assessment, exposure assessment, and risk characterization. Most recently, in the “Silver Book,” *Science and Decisions: Advancing Risk Assessment*, an NRC committee extended the framework of the Red Book in an effort to make risk assessments more useful for decision-making (NRC 2009). Those and other reports have consistently highlighted the necessity for comprehensive assessment of evidence and characterization of uncertainty and variability, and the Silver Book emphasizes assessment of uncertainty and variability appropriate to the decision to be made.

Science and Decisions: Advancing Risk Assessment made several recommendations directly relevant to developing IRIS assessments, including the draft formaldehyde assessment. First, it called for the development of guidance related to the handling of uncertainty and variability, that is, clear definitions and methods. Second, it urged a unified dose-response assessment framework for chemicals that would link understanding of disease processes, modes of action, and human heterogeneity among cancer and noncancer outcomes. Thus, it suggested an expansion of cancer dose-response assessments to reflect variability and uncertainty more fully and for noncancer dose-response assessments to reflect analysis of the probability of adverse responses at particular exposures. Although that is an ambitious undertaking, steps toward a unifying framework would benefit future IRIS assessments. Third, the Silver Book recommended that EPA assess its capacity for risk assessment and take steps to ensure that it is able to carry out its challenging risk-assessment agenda. For some IRIS assessments, EPA appears to have difficulty in assembling the needed multidisciplinary teams.

The committee recognizes that EPA has initiated a plan to revise the overall IRIS process and issued a memorandum that provided a brief description of the steps (EPA 2009a). Figure 7-1 illustrates the steps outlined in that memorandum. The committee is concerned that little information is provided on what it sees as the most critical step, that is, completion of a draft IRIS assessment. In the flow diagram, six steps are devoted to the review process, and thus the focus of the revision appears to be on the steps after the assessment has been generated. Although EPA may be revising its approaches for completing the draft assessment (Step 1 in Figure 7-1), the committee could not locate any other information on the revision of the IRIS process. Therefore, the committee offers some suggestions on the development process.

In providing guidance on revisions of the IRIS development process (that is, Step 1 as illustrated in Figure 7-1), the committee begins with a discussion of the current state of science regarding reviews of evidence and cites several examples that provide potential models for IRIS assessments. The

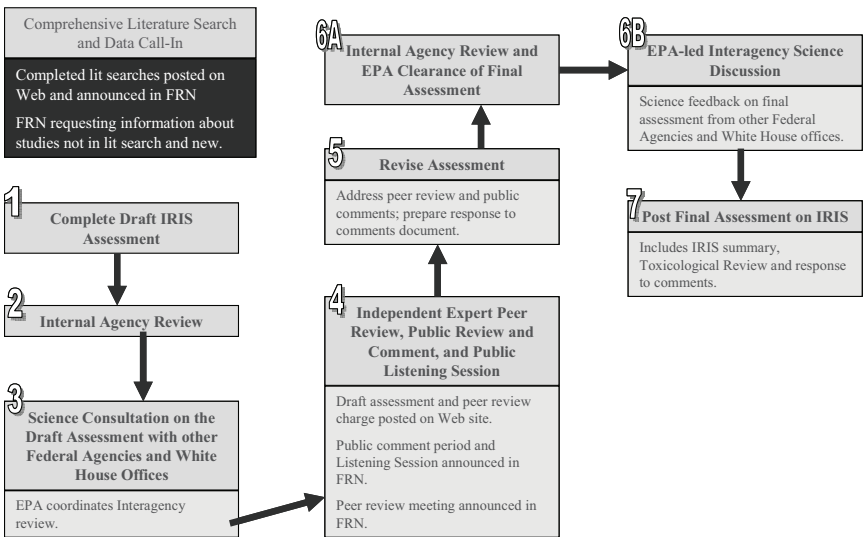


FIGURE 7-1 New IRIS assessment process. Abbreviations: FRN, Federal Register Notice; IRIS, Integrated Risk Information System; and EPA, Environmental Protection Agency. Source: EPA 2009a.

committee also describes the approach now followed in reviewing and synthesizing evidence related to the National Ambient Air Quality Standards (NAAQSs), a process that has been modified over the last 2 years. It is provided as an informative example of how the agency was able to revise an entrenched process in a relatively short time, not as an example of a specific process that should be adopted for the IRIS process. Finally, the committee offers some suggestions for improving the IRIS development process, providing a “roadmap” of the specific items for consideration.

An Overview of the Development of the Draft IRIS Assessment

In Chapter 2, the committee provided its own diagram (Figure 2-1) describing the steps used to generate the draft IRIS assessment. For the purpose of offering committee comments on ways to improve those steps, that figure has been expanded to indicate the key outcomes at each step (Figure 7-2). For each of the steps, the figure identifies the key questions addressed in the process. At the broadest level, the steps include systematic review of evidence, hazard identification using a weight-of-evidence approach, and dose-response assessment.

The systematic review process is undertaken to identify all relevant literature on the agent of interest, to evaluate the identified studies, and possibly to

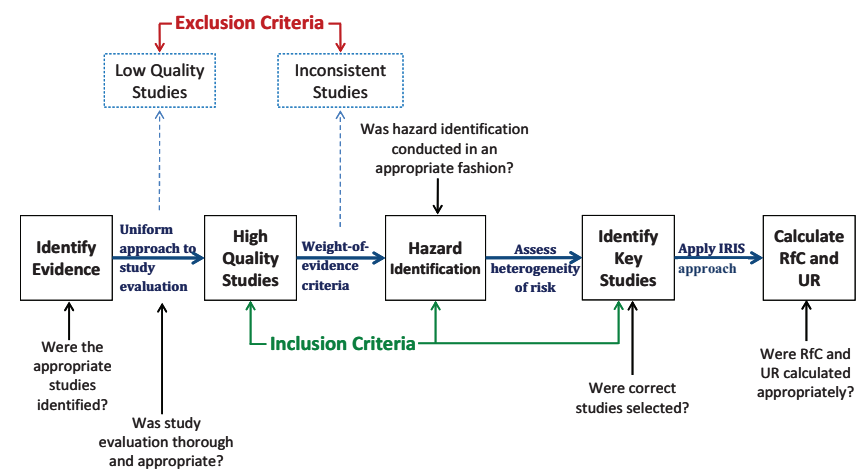


FIGURE 7-2 Elements of the key steps in the development of a draft IRIS assessment. Abbreviations: IRIS, Integrated Risk Information System; RfC, reference concentration; and UR, unit risk.

provide a qualitative or quantitative synthesis of the literature. Chapter 1 of the draft IRIS assessment of formaldehyde provides a brief general description of the process followed by EPA, including the approach to searching the literature. However, neither Chapter 1 nor other chapters of the draft provide a sufficiently detailed description of the approach taken in evaluating individual studies. In discussing particular epidemiologic studies, a systematic approach to study evaluation is not provided. Consequently, some of the key methodologic points are inconsistently mentioned, such as information bias and confounding.

For hazard identification, the general guidance is also found in Chapter 1 of the draft IRIS assessment. The approach to conducting hazard identification is critical for the integrity of the IRIS process. The various guidelines cited in Chapter 1 provide a general indication of the approach to be taken to hazard identification but do not offer a clear template for carrying it out. For the formaldehyde assessment, hazard identification is particularly challenging because the outcomes include cancer and multiple noncancer outcomes. The various EPA guidelines themselves have not been harmonized, and they provide only general guidance. Ultimately, the quality of the studies reviewed and the strength of evidence provided by the studies for deriving RfCs and unit risks need to be clearly presented. More formulaic approaches are followed for calculation of RfCs and unit risks. The key issue is whether the calculations were conducted appropriately and according to accepted assessment procedures.

Brief Review of Established Best Practices

The following sections highlight some best practices of current approaches to evidence-based reviews, hazard identification, and dose-response assessment that could provide EPA guidance if it decides to address some of the fundamental issues identified by the committee. The discussion is meant not to be comprehensive or to provide all perspectives on the topics but simply to highlight some important aspects of the approaches. The committee recognizes that some of the concepts and approaches discussed below are elementary and are addressed in some of EPA's guidelines. However, the current state of the formaldehyde draft IRIS assessment suggests that there might be a problem with the practical implementation of the guidelines in completing the IRIS assessments. Therefore, the committee highlights aspects that it finds most critical.

Current Approaches to Evidence-Based Reviews

Public-health decision-making has a long history of using comprehensive reviews as the foundation for evaluating evidence and selecting policy options. The landmark 1964 report of the U.S. surgeon general on tobacco and disease is exemplary (DHEW 1964). It used a transparent method that involved a critical survey of all relevant literature by a neutral panel of experts and an explicit framework for assessing the strength of evidence for causation that was equivalent to hazard identification (Table 7-1).

The tradition of comprehensive, evidence-based reviews has been continued in the surgeon general's reports. The 2004 surgeon general's report, which marked the 40th anniversary of the first report, highlighted the approach for causal inference used in previous reports and provided an updated and standardized four-level system for describing strength of evidence (DHHS 2004) (Table 7-2).

The same systematic approaches have become fundamental in many fields of clinical medicine and public health. The paradigm of "evidence-based medicine" involves the systematic review of evidence as the basis of guidelines. The international Cochrane Collaboration engages thousands of researchers and clinicians throughout the world to carry out reviews. In the United States, the Agency for Healthcare Research and Quality supports 14 evidence-based practice centers to conduct reviews related to healthcare.

There are also numerous reports from NRC committees and the Institute of Medicine (IOM) that exemplify the use of systematic reviews in evaluating evidence. Examples include reviews of the possible adverse responses associated with Agent Orange, vaccines, asbestos, arsenic in drinking water, and secondhand smoke. A 2008 IOM report, *Improving the Presumptive Disability Decision-Making Process for Veterans*, proposed a comprehensive new scheme for

TABLE 7-1 Criteria for Determining Causality

Criterion	Definition
Consistency	Persistent association among different studies in different populations
Strength of association	Magnitude of the association
Specificity	Linkage of specific exposure to specific outcome
Temporality	Exposure comes before effect
Coherence, plausibility, analogy	Coherence of the various lines of evidence with a causal relationship
Biologic gradient	Presence of increasing effect with increasing exposure (dose-response relationship)
Experiment	Observations from “natural experiments,” such as cessation of exposure (for example, quitting smoking)

Source: DHHS 2004.

TABLE 7-2 Hierarchy for Classifying Strength of Causal Inferences on the Basis of Available Evidence

A. Evidence is <i>sufficient</i> to infer a causal relationship.
B. Evidence is <i>suggestive but not sufficient</i> to infer a causal relationship.
C. Evidence is <i>inadequate</i> to infer the presence or absence of a causal relationship (evidence that is sparse, of poor quality, or conflicting).
D. Evidence is <i>suggestive of no causal relationship</i> .

Source: DHHS 2004.

evaluating evidence that an exposure sustained in military service had contributed to disease (IOM 2008); the report offers relevant coverage of the practice of causal inference.

This brief and necessarily selective coverage of evidence reviews and evaluations shows that models are available that have proved successful in practice. They have several common elements: transparent and explicitly documented methods, consistent and critical evaluation of all relevant literature, application of a standardized approach for grading the strength of evidence, and clear and consistent summative language. Finally, highlighting features and limitations of the studies for use in quantitative assessments seems especially important for IRIS literature reviews.

A state-of-the-art literature review is essential for ensuring that the process of gathering evidence is comprehensive, transparent, and balanced. The committee suggests that EPA develop a detailed search strategy with search terms related to the specific questions that are addressed by the literature review. The yield of articles from searches can best be displayed graphically, documenting how initial search findings are narrowed to the articles in the final review selection on the basis of inclusion and exclusion criteria. Figure 7-3 provides an example of the selection process in a systematic review of a drug for lung disease. The progression from the initial 3,153 identified articles to the 11 reviewed is transparent. Although this example comes from an epidemiologic meta-analysis, a similar transparent process in which search terms, databases, and resources are listed and study selection is carefully tracked may be useful at all stages of the development of the IRIS assessment.

After studies are identified for review, the next step is to summarize the details and findings in evidence tables. Typically, such tables provide a link to the references, details of the study populations and methods, and key findings. They are prepared in a rigorous fashion with quality-assurance measures, such as using multiple abstractors (at least for a sample) and checking all numbers abstracted. If prepared correctly, the tables eliminate the need for long descriptions of studies and result in shorter text. Some draft IRIS assessments have begun to use a tabular format for systematic and concise presentation of evidence, and the committee encourages EPA to refine and expand that format as it revises the formaldehyde draft IRIS assessment and begins work on others.

The methods and findings of the studies are then evaluated with a standardized approach. Templates are useful for this purpose to ensure uniformity of approach, particularly if multiple reviewers are involved. Such standardized approaches are applied whether the research is epidemiologic (observational), experimental (randomized clinical trials), or toxicologic (animal bioassays). For example, for an observational epidemiologic study, a template for evaluation should consider the following:

- Approach used to identify the study population and the potential for selection bias.
- Study population characteristics and the generalizability of findings to other populations.
- Approach used for exposure assessment and the potential for information bias, whether differential (nonrandom) or nondifferential (random).
- Approach used for outcome identification and any potential bias.
- Appropriateness of analytic methods used.
- Potential for confounding to have influenced the findings.
- Precision of estimates of effect.
- Availability of an exposure metric that is used to model the severity of adverse response associated with a gradient of exposures.

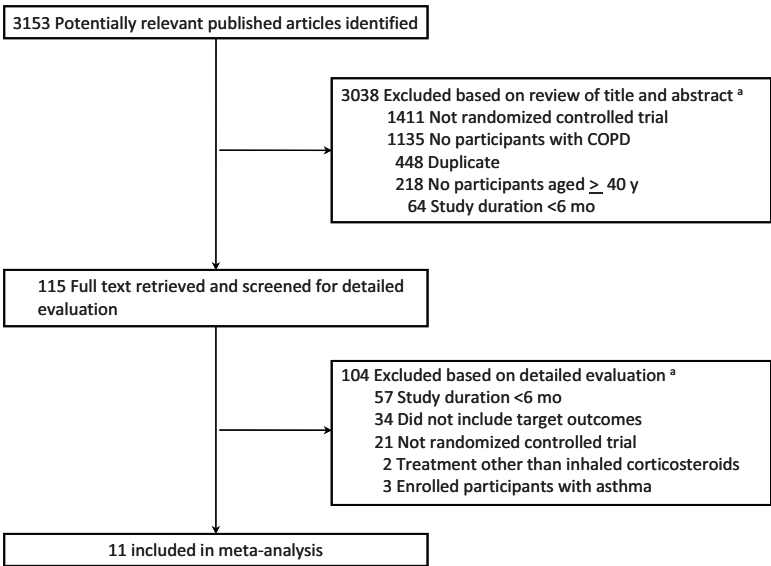


FIGURE 7-3 Example of an article-selection process. ^aArticles could be excluded for more than one reason; therefore, summed exclusions exceed total. Abbreviation: COPD, chronic obstructive pulmonary disease. Source: Drummond et al. 2008. Reprinted with permission; copyright 2008, American Medical Association.

Similarly, a template for evaluation of a toxicology study in laboratory animals should consider the species and sex of animals studied, dosing information (dose spacing, dose duration, and route of exposure), end points considered, and the relevance of the end points to human end points of concern.

Current Approaches to Hazard Identification

Hazard identification involves answering the question, Does the agent cause the adverse effect? (NRC 1983, 2009). Numerous approaches have been used for this purpose, and there is an extensive literature on causal inference, both on its philosophic underpinnings and on methods for evaluating the strength of evidence of causation. All approaches have in common a systematic identification of relevant evidence, criteria for evaluating the strength of evidence, and language for describing the strength of evidence of causation. The topic of causal inference and its role in decision-making was recently covered in the 2008 IOM report on evaluation of the presumptive decision-making process noted above. The 2004 report of the U.S. surgeon general on smoking and health (DHHS 2004) provided an updated review of the methods used in that series of reports.

The review approach for hazard identification embodies the elements described above and uses the criteria for evidence evaluation that have their origins in the 1964 report of the U.S. surgeon general (DHEW 1964) and the writings of Austin Bradford Hill, commonly known as the Hill criteria (see Table 7-1; Hill 1965). The criteria are not rigid and are not applied in a check-list manner; in fact, none is required for inferring a causal relationship, except for temporality inasmuch as exposure to the causal agent must precede the associated effect. The conclusion of causal inference is a clear statement on the strength of evidence of causation. For the purpose of hazard identification, such statements should follow a standardized classification to avoid ambiguity and to ensure comparability among different agents and outcomes.

Beyond the surgeon general's reports used here as an example, there are numerous examples of systematic approaches to hazard identification, including the monographs on carcinogenicity of the International Agency for Research on Cancer and the National Toxicology Program.¹ They have the same elements of systematic gathering and review of all lines of evidence and classification of the strength of evidence in a uniform and hierarchic structure.

Current Approaches to Dose-Response Assessment

The topic of dose-response assessment was covered in *Science and Decisions* (NRC 2009), which reviewed the current paradigm and called for a unified framework, bringing commonality to approaches for cancer and noncancer end points. That report also provides guidance on enhancing methods used to characterize uncertainty and variability. The present committee supports those recommendations but offers additional suggestions on the complementary coverage of the use of meta-analysis and pooled analysis in dose-response assessment.

IRIS assessments should address the following critical questions: Which studies should be included for derivation of reference values for noncancer outcomes and unit risks for cancer outcomes? Which dose-response models should be used for deriving those values? The latter question is related to model uncertainty in quantitative risk assessment and is not addressed here in this report. The former question is related to a fundamental issue of filtering the literature to identify the studies that provide the best dose-response information. A related question arises about how to combine information among studies because multiple studies may provide sufficient dose-response data. For this section, the committee assumes that the previously described evidence-based review has identified studies with adequate dose-response information to support some quantification of risk associated with exposure.

As suggested above, it would be unusual for a single study to trump all other studies providing information for setting reference values and unit risks. The combination of the analysis outcomes of different studies falls under the

¹See <http://monographs.iarc.fr/index.php> and <http://ntp.niehs.nih.gov/>.

general description of meta-analysis (Normand 1999). The combination and synthesis of results of different studies appears central to an IRIS assessment, but such analyses require careful framing.

Stroup and colleagues (2000) provide a summary of recommendations for reporting meta-analyses of epidemiologic studies. Their proposal includes a table with a proposed check list that has broad categories for reporting, including background (such as problem definition and study population), search strategy (such as searchers, databases, and registries used), methods, results (such as graphic and tabular summaries, study description, and statistical uncertainty), discussion (such as bias and quality of included studies), and conclusion (such as generalization of conclusions and alternative explanations). Their recommendations on methods warrant specific consideration with reference to the development of an IRIS assessment, particularly those on evaluation and assessment of study relevance, rationale for selection and coding of studies, confounding, study quality, heterogeneity, and statistical methods. For the latter, key issues include the selection of models, the clarity with which findings are presented, and the availability of sufficient details to facilitate replication.

In combining study information, it is important that studies provide information on the same quantitative outcome, are conducted under similar conditions, and are of similar quality. If studies are of different quality, this might be addressed by weighting.

The simplest form of combining study information involves the aggregation of *p* values among a set of independent studies of the same null hypothesis. That simple approach might have appeal for establishing the relationship between some risk factor and an adverse outcome, but it is not useful for establishing exposure levels for a hazard. Thus, effect-size estimation among studies is usually of more interest for risk-estimation purposes and causality assessment. In this situation, a given effect is estimated for each study, and a combined estimate is obtained as a weighted average of study-specific effects in which the weights are inversely related to the precision associated with the estimation of each study-specific effect.

The question is whether EPA should routinely conduct meta-analysis for its IRIS assessments. Implicitly, the development of an IRIS assessment involves many of the steps associated with meta-analysis, including the collection and assessment of background literature. Assuming the availability of independent studies of the same end point and a comprehensive and unbiased inclusion of studies, questions addressed by a meta-analysis may be of great interest. Is there evidence of a homogeneous effect among studies? If not, can one understand the source of heterogeneity? If it is determined that a combined estimate is of interest (for example, an estimate of lifetime cancer risk based on combining study-specific estimates of this risk), a weighted estimate might be derived and reported.

Case Study: Revision of the Approach to Evidence Review and Risk Assessment for National Ambient Air Quality Standards

Approaches to evidence review and risk assessment vary within EPA. The recently revised approach used for NAAQSs offers an example that is particularly relevant because it represents a major change in an approach taken by one group in the National Center for Environmental Assessment. (EPA 2009b, 2010a,b)

Under Section 109 of the Clean Air Act, EPA is required to consider revisions of the NAAQSs for specified criteria air pollutants—currently particulate matter (PM), ozone, nitrogen dioxide, sulfur dioxide, carbon monoxide, and lead—every 5 years. Through 2009, the process for revision involved the development of two related documents that were both reviewed by the Clean Air Scientific Advisory Committee (CASAC) and made available for public comment. The first, the criteria document, was an encyclopedic compilation, sometimes several thousand pages long, of most scientific publications on the criteria pollutant that had been published since the previous review. Multiple authors contributed to the document, and there was generally little synthesis of the evidence, which was not accomplished in a systematic manner.

The other document was referred to as the staff paper. It was written by a different team in the Office of Air Quality Policy and Standards, and it identified the key scientific advances in the criteria document that were relevant to revising the NAAQSs. In the context of those advances, it offered the array of policy options around retaining or revising the NAAQSs that could be justified by recent research evidence. The linkages between the criteria document and the staff paper were general and not transparent.

The identified limitations of the process led to a proposal for its revision, and it took 2 years to complete the changes in the process. The new process replaces the criteria document with an integrated science assessment and a staff paper that includes a policy assessment. For the one pollutant, PM, that has nearly completed the full sequence, a risk and exposure analysis was also included.

The new documents address limitations of those used previously. The integrated science assessment is an evidence-based review that targets new studies as before. However, review methods are explicitly stated, and studies are reviewed in an informative and purposeful manner rather than in encyclopedic fashion. A main purpose of the integrated science assessment is to assess whether adverse health effects are causally linked to the pollutant under review. The integrated science assessment offers a five-category grading of strength of evidence on each outcome and follows the general weight-of-evidence approaches long used in public health. The intent is to base the risk and exposure analysis on effects for which causality is inferred or those at lower levels if they have particular public-health significance. The risk and exposure analysis brings

together the quantitative information on risk and exposure and provides estimates of the current burden of attributable morbidity and mortality and the estimates of avoidable and residual morbidity and mortality under various scenarios of changes in the NAAQS. Standard descriptors for uncertainty are now in place.

The policy assessment develops policy options on the basis of the findings of the integrated science assessment and the risk and exposure analysis. The policy assessment for the PM NAAQS is framed around a series of policy-relevant questions, such as, Does the available scientific evidence, as reflected in the integrated science assessment, support or call into question the adequacy of the protection afforded by the current 24-hr PM₁₀ standard against effects associated with exposures to thoracic coarse particles? Evidence-based answers to the questions are provided with a reasonably standardized terminology for uncertainty.

For the most recent reassessment of the PM NAAQS, EPA staff and CASAC found the process to be effective; it led to greater transparency in evidence review and development of policy options than the prior process (Samet 2010). As noted above, the present committee sees the revision of the NAAQS review process as a useful example of how the agency was able to revise an entrenched process in a relatively short time.

Reframing the Development of the IRIS Assessment

The committee was given the broad charge of reviewing the formaldehyde draft IRIS assessment and also asked to consider some specific questions. In addressing those questions, the committee found, as documented in Chapter 2, that some problems with the draft arose because of the processes and methods used to develop the assessment. Other committees have noted some of the same problems. Accordingly, the committee suggests here steps that EPA could take to improve IRIS assessment through the implementation of methods that would better reflect current practices. The committee offers a roadmap for changes in the development process if EPA concludes that such changes are needed. The term *roadmap* is used because the topics that need to be addressed are set out, but detailed guidance is not provided because that is seen as beyond the committee's charge. The committee's discussion of a reframing of the IRIS development process is based on its generic representation provided in Figure 7-2. The committee recognizes that the changes suggested would involve a multiyear process and extensive effort by the staff of the National Center for Environmental Assessment and input and review by the EPA Science Advisory Board and others. The recent revision of the NAAQS review process provides an example of an overhauling of an EPA evidence-review and risk-assessment process that took about 2 years.

In the judgment of the present and past committees, consideration needs to be given to how each step of the process could be improved and gains made in transparency and efficiency. Models for conducting IRIS reviews more effectively and efficiently are available. For each of the various components (Figure 7-2), methods have been developed, and there are exemplary approaches in assessments carried out elsewhere in EPA and by other organizations. In addition, there are relevant examples of evidence-based algorithms that EPA could draw on. Guidelines and protocols for the conduct of evidence-based reviews are available, as are guidelines for inference as to the strength of evidence of association and causation. Thus, EPA may be able to make changes in the assessment process relatively quickly by drawing on appropriate experts and selecting and adapting existing approaches.

One major, overarching issue is the use of weight of evidence in hazard identification. The committee recognizes that the terminology is embedded in various EPA guidelines (see Appendix B) and has proved useful. The determination of weight of evidence relies heavily on expert judgment. As called for by others, EPA might direct effort at better understanding how weight-of-evidence determinations are made with a goal of improving the process (White et al. 2009).

The committee highlights below what it considers critical for the development of a scientifically sound IRIS assessment. Although many elements are basic and have been addressed in the numerous EPA guidelines, implementation does not appear to be systematic or uniform in the development of the IRIS assessments.

General Guidance for the Overall Process

- Elaborate an overall, documented, and quality-controlled process for IRIS assessments.
- Ensure standardization of review and evaluation approaches among contributors and teams of contributors; for example, include standard approaches for reviews of various types of studies to ensure uniformity.
- Assess disciplinary structure of teams needed to conduct the assessments.

Evidence Identification: Literature Collection and Collation Phase

- Select outcomes on the basis of available evidence and understanding of mode of action.
- Establish standard protocols for evidence identification.
- Develop a template for description of the search approach.
- Use a database, such as the Health and Environmental Research Online (HERO) database, to capture study information and relevant quantitative data.

Evidence Evaluation: Hazard Identification and Dose-Response Modeling

- Standardize the presentation of reviewed studies in tabular or graphic form to capture the key dimensions of study characteristics, weight of evidence, and utility as a basis for deriving reference values and unit risks.
 - Develop templates for evidence tables, forest plots, or other displays.
 - Establish protocols for review of major types of studies, such as epidemiologic and bioassay.

Weight-of-Evidence Evaluation: Synthesis of Evidence for Hazard Identification

- Review use of existing weight-of-evidence guidelines.
- Standardize approach to using weight-of-evidence guidelines.
- Conduct agency workshops on approaches to implementing weight-of-evidence guidelines.
 - Develop uniform language to describe strength of evidence on noncancer effects.
 - Expand and harmonize the approach for characterizing uncertainty and variability.
 - To the extent possible, unify consideration of outcomes around common modes of action rather than considering multiple outcomes separately.

Selection of Studies for Derivation of Reference Values and Unit Risks

- Establish clear guidelines for study selection.
 - Balance strengths and weaknesses.
 - Weigh human vs experimental evidence.
 - Determine whether combining estimates among studies is warranted.

Calculation of Reference Values and Unit Risks

- Describe and justify assumptions and models used. This step includes review of dosimetry models and the implications of the models for uncertainty factors; determination of appropriate points of departure (such as benchmark dose, no-observed-adverse-effect level, and lowest observed-adverse-effect level), and assessment of the analyses that underlie the points of departure.
 - Provide explanation of the risk-estimation modeling processes (for example, a statistical or biologic model fit to the data) that are used to develop a unit risk estimate.

- Assess the sensitivity of derived estimates to model assumptions and end points selected. This step should include appropriate tabular and graphic displays to illustrate the range of the estimates and the effect of uncertainty factors on the estimates.
- Provide adequate documentation for conclusions and estimation of reference values and unit risks. As noted by the committee throughout the present report, sufficient support for conclusions in the formaldehyde draft IRIS assessment is often lacking. Given that the development of specific IRIS assessments and their conclusions are of interest to many stakeholders, it is important that they provide sufficient references and supporting documentation for their conclusions. Detailed appendixes, which might be made available only electronically, should be provided when appropriate.

REFERENCES

- DHEW (U.S. Department of Health Education and Welfare). 1964. Smoking and Health. Report of the Advisory Committee to the Surgeon General. Public Health Service Publication No. 1103. Washington, DC: U.S. Government Printing Office [online]. Available: http://profiles.nlm.nih.gov/NN/B/B/M/Q/_/nnbbmq.pdf [accessed Feb. 1, 2011].
- DHHS (U.S. Department of Health and Human Services). 2004. The Health Consequences of Smoking: A Report of the Surgeon General. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta, GA [online]. Available: http://www.cdc.gov/tobacco/data_statistics/sgr/2004/complete_report/index.htm [accessed Nov. 22, 2010].
- Drummond, M.B., E.C. Dasenbrook, M.W. Pitz, D.J. Murphy, and E. Fan. 2008. Inhaled corticosteroids in patients with stable chronic obstructive pulmonary disease: A systematic review and meta-analysis. *JAMA*. 300(20):2407-2416.
- EPA (U.S. Environmental Protection Agency). 2009a. New Process for Development of Integrated Risk Information System Health Assessments. Memorandum to Assistant Administrators, General Counsel, Inspector General, Chief Financial Officer, Chief of Staff, Associate Administrators, and Regional Administrators, from Lisa P. Jackson, the Administrator, U.S. Environmental Protection Agency, Washington, DC. May 21, 2009 [online]. Available: http://www.epa.gov/iris/pdfs/IRIS_PROCESS_MEMO.5.21.09.PDF [accessed Nov. 23, 2010].
- EPA (U.S. Environmental Protection Agency). 2009b. Integrated Science Assessment for Particulate Matter (Final Report). EPA/600/R-08/139F. National Center for Environmental Assessment-RTP Division, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC. December 2009 [online]. Available: <http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=216546> [accessed March 2, 2011].
- EPA (U.S. Environmental Protection Agency). 2010a. Quantitative Health Risk Assessment for Particulate Matter (Final Report). EPA-452/R-10-005. Office of Air Quality Planning and Standards, Office of Air and Radiation, U.S. Environmental Protection Agency, Research Triangle Park, NC. June 2010.

- [online]. Available: http://www.epa.gov/ttn/naaqs/standards/pm/data/PM_RA_FINAL_June_2010.pdf [accessed March 2, 2011].
- EPA (U.S. Environmental Protection Agency). 2010b. Policy Assessment for the Review of the Particulate Matter National Ambient Air Quality Standards (Second External Review Draft). EPA 452/P-10-007. Office of Air Quality Planning and Standards, Office of Air and Radiation, U.S. Environmental Protection Agency, Research Triangle Park, NC. June 2010 [online]. Available: <http://www.epa.gov/ttnnaaqs/standards/pm/data/20100630seconddraftmpa.pdf> [accessed March 2, 2011].
- Hill, A.B. 1965. The environment and disease: Association or causation? *Proc. R. Soc. Med.* 58:295-300
- IOM (Institute of Medicine). 2008. Improving the Presumptive Disability Decision-Making Process for Veterans. Washington, DC: National Academies Press.
- NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. Washington, DC: National Academy Press.
- NRC (National Research Council). 2009. Science and Decisions: Advancing Risk Assessment. Washington, DC: National Academies Press.
- Normand, S.L. 1999. Meta-analysis: Formulating, evaluating, combining, and reporting. *Stat. Med.* 18(3): 321-359.
- Samet, J.M. 2010. CASAC Review of Policy Assessment for the Review of the PM NAAQS - Second External Review Draft (June 2010). EPA-CASAC-10-015. Letter to Lisa P. Jackson, Administrator, from Jonathan M. Samet, Clean Air Scientific Advisory Committee, Office of Administrator, Science Advisory Board, U.S. Environmental Protection Agency, Washington, DC. September 10, 2010 [online]. Available: [http://yosemite.epa.gov/sab/sabproduct.nsf/CCF9F4C0500C500F8525779D0073C593/\\$File/EPA-CASAC-10-015-unsigned.pdf](http://yosemite.epa.gov/sab/sabproduct.nsf/CCF9F4C0500C500F8525779D0073C593/$File/EPA-CASAC-10-015-unsigned.pdf) [accessed Nov. 22, 2010].
- Stroup, D.F., J.A. Berlin, S.C. Morton, I. Olkin, G.D. Williamson, D. Rennie, D. Moher, B.J. Becker, T.A. Sipe, and S.B. Thacker. 2000. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA.* 283(15):2008-2012.
- White, R.H., I. Cote, L. Zeise, M. Fox, F. Dominici, T.A. Burke, P.D. White, D.B. Hattis, and J.M. Samet. 2009. State-of-the-science workshop report: Issues and approaches in low-dose-response extrapolation for environmental health risk assessment. *Environ. Health Perspect.* 117(2):283-287.